

数智技术驱动下的学术资源库建设新范式 研究：从多元数据集成到知识服务的转型

汇报人：高振
2025年 5月 16日



山东大学
SHANDONG UNIVERSITY

C 目 录 CONTENTS

一、引言

二、数智技术赋能学术资源库转型的理论基础

三、多元数据集成的实现路径

四、知识服务体系的构建

五、开放获取与资源共享机制

六、实践路径与未来发展方向

七、结语



山东大学
SHANDONG UNIVERSITY

一、引言



研究背景

现代科研需求

现代科研需求日益复杂化，强调精细化数据、跨学科整合和智能化服务。通过调研我们发现，超过85%的科研人员希望学术数据库提供跨学科的数据整合和深入的数据分析等智能服务。

01

传统学术资源库的局限

传统学术资源库主要以文献存储为核心功能，个性化服务能力不足，难以满足当前科研的多样化需求。根据调研发现，传统资源库中超过65%的文献未被有效利用，造成了较大的资源浪费。

02

03

数智技术带来的新机遇

人工智能、知识图谱和自然语言处理等技术开始应用于学术资源推荐、学术数据分析，并取得了不错的效果，为资源库提供了智能化升级的可能性。



研究目标与核心问题

研究目标

基于上述背景，我们研究的目标是设计一个支持多源数据集成、定制化资源管理和个性化知识服务的学术资源库建设方案。我们希望通过数智技术的创新应用，使学术资源库从传统的“存储型”转变为“服务型”。

目标

核心问题

如何通过数智技术实现学术资源库由“存储”向“知识服务”的转型？也就是说，我们要找到技术与资源库建设的最佳结合点，以满足现代科研人员的实际需求。

问题

二、数智技术赋能学术资源库转型的 理论基础



关键技术支撑



人工智能

人工智能尤其是机器学习和深度学习技术，能够对海量学术数据进行自动分类与智能检索，提供高效的个性化推荐服务，显著提升用户的使用体验。



大数据

大数据技术通过分布式存储和计算框架，能够高效地集成和管理来自不同渠道的多源数据，应用于用户行为分析与学术趋势预测中，确保了数据处理能力的可扩展性和高效性。



知识图谱

知识图谱技术通过构建学术资源间的语义关联网络，大幅提升了跨学科知识发现能力，支持智能推荐和知识可视化等应用。



自然语言处理

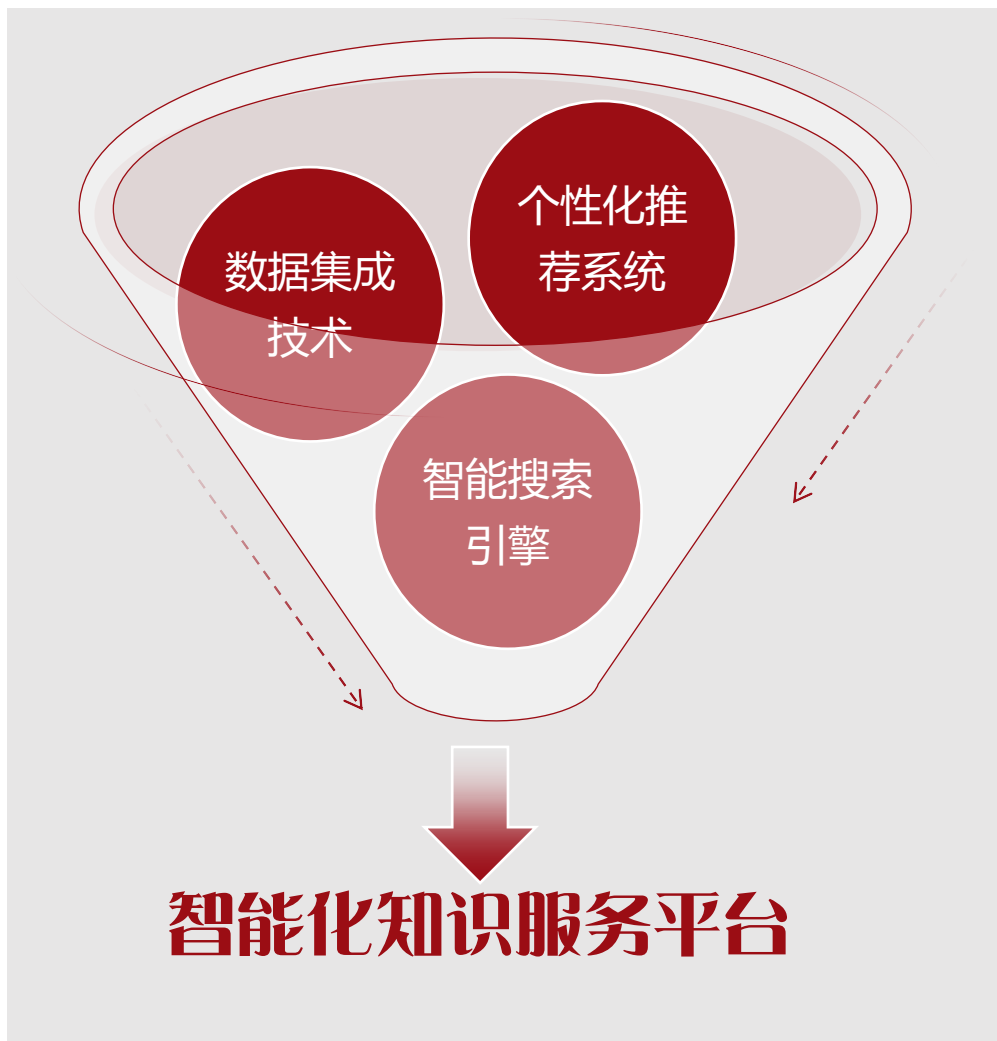
自然语言处理技术增强了资源库对学术文本的理解与处理能力，可实现精准语义检索、自动摘要和智能问答等服务。上述数智技术共同构成了我们理论基础的技术支撑部分。



关键技术支撑

技术类别	主要技术	应用场景	具体应用	对学术资源库的影响
人工智能（AI）	机器学习、深度学习	自动化资源分类、智能检索、智能推荐	自动分类与标注、语义检索、个性化推荐	提高资源管理效率，提升检索和推荐的精准度
大数据技术	分布式存储、计算框架	多源数据集成、用户行为分析、趋势分析	多源数据集成、用户兴趣建模、学术趋势预测	提升数据处理能力，支持大规模数据集成与分析
自然语言处理（NLP）	语义分析、文本生成	语义检索、自动摘要、智能问答	语义检索、自动摘要生成、智能问答系统	提升语义理解能力，优化用户检索体验
知识图谱	语义关联、可视化	跨学科知识关联、智能推荐、可视化	跨学科知识关联、智能推荐、学术资源可视化	促进跨学科知识发现，提升知识推荐和可视化能力

理论模型



“数据集成+知识服务”模型

基于这些技术支撑，课题组建立了学术资源库转型的“双支撑”理论框架，即“数据集成+知识服务”模型：一方面，通过关联数据、语义计算和知识图谱等方法，实现多源异构学术资源的语义整合；另一方面，基于集成的数据，运用协同过滤、深度学习等个性化推荐模型，将数据价值转化为知识服务能力，从而推动学术资源库从“数据存储”向“知识赋能”的跃迁。

相关理论的耦合问题



01

多源数据的语义对齐目前仍主要依赖领域专家的深度参与，自动化程度有限。



02

如果深度学习模型缺乏可解释性，可能会引发用户对推荐结果的信任危机。



03

开放共享与隐私保护的平衡问题尚未完全解决，需通过技术手段在相关模型本地化训练中保护用户数据。

数据集成与知识服务的理论融合，体现在“数据—知识—服务”的闭环链条中。当然，在实践过程中将数据集成与知识服务融合为闭环模型仍面临着多重问题。



山东大学
SHANDONG UNIVERSITY

三、多元数据集成的实现路径

核心挑战

在学术资源库中，我们需要整合来自多种来源的异构数据，包括文献、实验数据、计算模型、开放数据集等。由于数据类型丰富、格式多样，彼此缺乏统一标准，多源数据格式的不统一和语义差异成为多源数据集成的核心难题。



多元数据

多元数据包括文献、实验数据、计算模型等，类型丰富、格式多样。



核心问题

如何将不同来源、不同格式的庞杂数据进行有效融合和统一管理？

多源异构数据融合



标准化接口设计

标准化数据接口设计，通过统一协议消除异构系统之间的交互障碍。



语义标注与知识图谱构建

利用语义标注与构建知识图谱使数据具有机器可读的语义属性，推动数据从简单互联转向深层次的知识关联。



数据处理流程

应用数据清洗、融合与关联技术确保数据质量。

通过以上技术手段的协同应用，可以有效的集成多源异构数据，提高学术资源库的数据融合能力。



标准化数据管理与更新



机器学习自动分类

利用机器学习自动对学术数据进行分类。



主题标签生成

引入主题标签生成算法为学术数据添加标签。



数据更新的自动检测与动态管理

建立数据自动更新检测与动态管理机制保证资源库数据的实时性与准确性。

通过这些技术手段的综合应用，使得在数据规模指数级增长和科研需求快速变化的情况下，资源库仍然能够实时、精准地集成和提供学术资源。

数据集成管理体系

通过多源异构数据融合与统一管理新技术的协同应用，构建了学术资源库“感知-处理-响应”的闭环数据集成管理体系。



四、知识服务体系的构建

智能搜索引擎

01

自然语言处理技术优化语义搜索，使用户能更准确地获取所需信息。语义搜索可理解用户查询意图，提供更精准的搜索结果。



自然语言处理

02

语义理解优化技术可提高搜索引擎的准确性和效率。通过语义理解，搜索引擎可识别用户的真实需求，提供更相关的结果。



语义理解优化

03

随着多模态学习（如图文跨模态检索）与因果推理技术的发展，语义搜索引擎有望进一步理解学术资源的深层逻辑链条，为科研决策提供智能化支持。

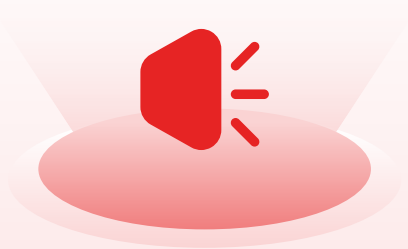


多模态学习

智能搜索引擎通过自然语言处理和多模态学习技术准确理解并满足用户的具体检索需求。

个性化推荐系统

协同过滤算法



协同过滤算法通过分析用户行为和偏好，为用户提供个性化推荐。协同过滤可发现用户间的相似性，为用户提供更符合其兴趣的资源。

深度学习技术



深度学习技术可进一步优化个性化推荐，提高推荐的准确性和多样性。深度学习技术可学习用户复杂的兴趣模式，提供更精准的推荐。

用户行为建模与兴趣偏好分析



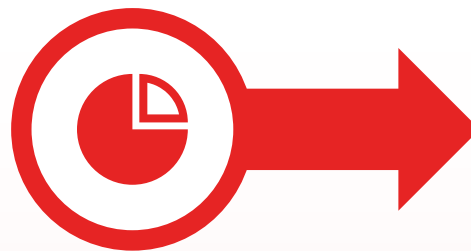
用户行为建模的核心在于从多源行为数据中提取动态兴趣表征。兴趣偏好分析可解决兴趣漂移与多兴趣分离问题。

个性化推荐系统则基于协同过滤和深度学习技术，分析用户行为数据提供个性化资源推荐。

智能问答与科研助手系统

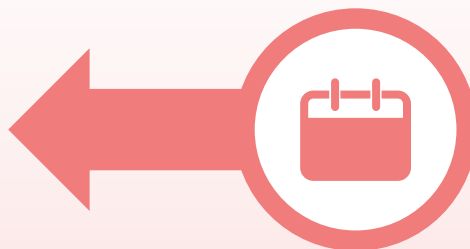
摘要自动生成

智能自动生成技术通过自然语言生成（NLG）模型，从长篇学术文献中提取核心内容，提升科研人员的文献处理效率。



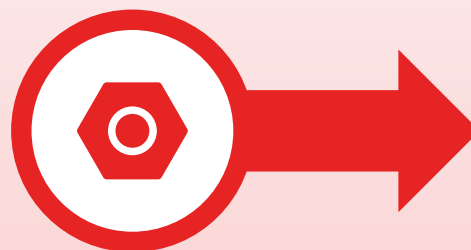
智能问答系统

AI驱动的智能问答系统通过预训练语言模型（如BERT、GPT-4）实现了对学术问题的深度语义理解与精准回答，突破传统关键词匹配的局限。



知识自动化机制

知识自动化机制是通过机器学习实现学术资源的动态治理与知识发现。其核心包括知识自动分类与关联和动态知识演化。



智能问答与科研助手系统依托预训练语言模型，实现对学术问题的深度理解与精准回答，并能够自动生成文献摘要，从而提高科研人员的工作效率。

知识服务体系



通过构建智能搜索、个性化推荐以及智能问答等知识服务系统模块，学术资源库能够为科研人员提供全方位的智能支持，真正实现由信息存储提供者向知识服务提供者的转变。



山东大学
SHANDONG UNIVERSITY

五、开放获取与资源共享机制



开放获取政策与标准

开放获取政策

开放获取政策的实施鼓励科研成果以开放形式发表和传播，使更多人能够平等地获取知识成果。

开放数据标准

开放数据标准的核心在于确保科研数据具备良好的可发现性、可互操作性和可重用性，使不同机构和研究者能够方便地查找、理解并利用这些数据。

资源共享协议

我们还需要制定明确的资源共享协议，从法律和技术层面规范学术资源的开放共享，在促进资源开放的同时保障作者权益和隐私安全，平衡开放与权益保护的关系。

科研资助机构和政府部门通过制定并推广开放获取政策与标准，体现了对知识公共性的高度重视。



标准化资源共享与调度机制

01

资源集中分配的智能 调度策略

利用动态优化算法实现资源的高效分配，最大程度地提高资源的利用效率。

02

隐私保护与数据 安全技术

通过隐私保护和数据安全技术，保障了数据共享过程中的安全性与合规性，推进全球范围内学术资源的开放共享。

通过完善开放获取政策、标准和共享机制，学术资源库能够更好地实现知识的开放流通，同时维护学术生态的公平和安全，促进科研成果的广泛传播。

六、学术资源库建设的新实践路径 与未来发展方向



学术资源库建设的新实践路径

01 云端资源库部署架构

采用云端部署、微服务与容器化架构，实现资源库的高可用性和良好扩展性，保障系统稳定运行。

03 资源库自动化运维

通过自动化运维工具，实现脚本化部署、实时监控，降低运维成本，快速响应故障。

05 资源更新与用户行为反馈闭环机制

建立事件驱动的数据更新机制和用户反馈闭环系统，确保资源库内容的实时更新和持续优化。

将AI、大数据、知识图谱技术进行深度融合，构建智能内核，提升跨模态数据检索的准确性和实时性。

02 AI+大数据+知识图谱的融合模式

构建多层次防御体系结合AI威胁预测模型，减少资源库的年均安全事件，进一步提升了资源库的安全防护能力。

04 安全防护

技术的适用问题与优化方向

现有技术的适用问题与优化方向

01

小样本与长尾问题

冷门学科（如古文字学）因数据稀疏性导致模型泛化能力不足。联邦学习与迁移学习（如Meta-Learning）或可通过跨机构协作缓解此问题，但需解决数据隐私与异构分布问题。统一检索。

02

多语言与多模态支持不足

现有系统对非英语资源的处理能力较弱，非洲土著语言文献、古汉语的解析准确率低于50%。多模态学习（如图文跨模态检索）虽在实验中取得进展，但其计算成本与实时性仍需优化。

03

隐私与安全的平衡难题

差分隐私与联邦学习虽能保护用户数据，但可能牺牲模型精度。未来需探索轻量化加密与边缘计算结合的新型隐私保护框架。



未来发展方向

01

技术深度融合

量子计算与边缘智能的引入有望突破算力瓶颈。例如，量子优化算法可加速大规模知识图谱的实时更新，而边缘节点的本地化处理将支持毫秒级响应。

03

人机协同的科研范式

AI将从“工具”升级为“协作者”。例如，因果推理引擎可自动生成实验改进建议，而生成式模型（如GPT-4）或能辅助论文撰写与同行评审，重塑科研工作流。

发展趋势与研究方向

开放科学的全球化协作

02

基于区块链的智能合约与去中心化存储将推动跨国资源的安全共享。OpenAIRE的实践经验表明，标准化协议与分布式治理模式是打破地域壁垒的关键。。

04

伦理与可持续发展

构建技术应用的伦理框架，防止算法偏见（如英语中心主义）与学术不端（如摘要生成的抄袭风险）。探索公益性与商业化的平衡模式，例如通过数据增值服务（如科研趋势分析报告）实现资源库的可持续运营。



山东大学
SHANDONG UNIVERSITY

七、结语

结语

通过本研究，我们课题组提出了一个由数智技术驱动的学术资源库建设新范式，并在数据集成、知识服务、开放共享等方面给出了创新性的解决方案。借助大数据、知识图谱、人工智能等技术手段，学术资源库不仅实现了从文献存储到智能知识服务的飞跃，也为科研人员提供了更高效、精准的学术支持。当然，学术资源库的建设与发展是一个持续的过程。展望未来，随着技术的不断进步和全球科研合作的深化，学术资源库将在知识共享和科研创新中发挥愈加重要的作用，成为推动开放科学和创新研究的重要基础设施。



为天下储人才 为国家图富强



山东大学
SHANDONG UNIVERSITY

谢谢

2025年5月16日